****

**A PROJECT RESEARCH ON**

**CUSTOMER SATISFACTION PREDICTION**

**USING MACHINE LEARNING**

**BY**

**OYEYEMI BOLUWATIFE JOSHUA**

**CSC/2018/1159**

**SUBMITTED TO THE**

**DEPARTMENT OF COMPUTER SCIENCE,**

**FACULTY OF SCIENCE,**

**FEDERAL UNIVERSITY OYE EKITI,**

**EKITI STATE, NIGERIA.**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF BACHELOR OF SCIENCE (B.Sc.)**

**IN COMPUTER SCIENCE.**

**NOVEMBER, 2023.**

**CERTIFICATION**

I hereby certify that this research work was carried out by **OYEYEMI BOLUWATIFE JOSHUA** with the matriculation number **CSC/2018/1159** under my supervision in the Department of Computer Science, Federal University Oye-Ekiti, Nigeria.

Mrs. Comfort Y. Daramola ……………………

Supervisor Signature/Date

Dr. Lawrence B. Adewole ……………………

Head of Department Signature/Date

i

**DECLARATION OF OWN WORK**

I certify that this project titled “Customer Satisfaction Prediction Using machine Learning” is entirely my own original work and that all sources utilized in its creation have been duly recognized and cited. I have been in charge of every aspect of this project's conception, design, and execution.

……………………

Oyeyemi Boluwatife Joshua

November, 2023.

ii

**DEDICATION**

This project is dedicated to God Almighty, the One who gives wisdom liberally without reproaching and also to my dynamic Parents who have made my education an utmost priority to them.

iii

**ACKNOWLEDGEMENT**

I acknowledge the Most High God, my Stone of Help for always coming through for me throughout my academic pursuit.

My profound gratitude to my amiable supervisor and mother, Mrs. Comfort YemisiDaramola for her support and encouragement since I started this research work. I cannot but appreciate my Head of Department, Dr. L.B. Adewole as well as all my esteemed lecturers.

A special thanks to my wonderful parents, Mr. &Chief (Mrs.) Emmanuel OyeniyiOyeyemi and sisters (Titilope, Temitope, Toyin and Testimony) for their financial support, constant love, prayers and encouragement throughout my degree program. Also, my appreciation goes to the family of Ven. & Mrs. B.S. Adedeji (FSA).

I would like to extend my heartfelt gratitude to every member of Anglican Students’ Fellowship, FUOYE, Oye Branch for making my stay in Oye worth it and for the spiritual growth, moral, academic and financial assistance and support when I needed them most. Also, I appreciate every member of St. John’s Anglican Church, Iye-Ekiti for their contributions in my success.

I would like to appreciate all my friends (Precious, Alaba, Stephen, Damilola, Christianah, Adesola, David and Boluwatife), housemates, mentees, mentors and reading partners for their dedication, collaboration, help, support and hard work. Their collective efforts and commitment to excellence have played a crucial role in achieving my goals in FUOYE.

iv

**ABSTRACT**

This research presents a machine learning approach to predicting customer satisfaction. The motivation for this research is to improve customer satisfaction by developing a reliable machine learning model. This study collects dataset from kaggle.com which is well known for housing different datasets that can be used for any data science project. The dataset was preprocessed using different preprocessing methods and the support vector machine and the Decision tree algorithms were used to train and test the model. The performance of the model was evaluated using a number of assessment parameters, including precision, recall, and F1 score. The Support Vector machine appeared to be the best after been trained with the dataset with an accuracy of 90% against the accuracy of 87% of the Decision Tree model. After checking the performance of previous studies, we realized that the models of this study performed better compared to other studies models. It is an improvement in predicting satisfaction behavior using Decision Tree and SVM models

*Keywords: Consumer Satisfaction, Machine Learning, Preprocessing, Feature Selection, Classification, Regression.*

v

**TABLE OF CONTENT**

CERTIFICATION…………………….………………………………………………………. i

[DECLARATION OF OWN WORK ………………..i](#_Toc121164355)i

DEDICATION…………………………………………………………………………………iii

[ACKNOWLEDGEMENT i](#_Toc121164356)v

ABSTRACT …………………………………………………………………………………. v

[TABLE OF CONTENT v](#_Toc121164358)i

**CHAPTER ONE: INTRODUCTION**

[1.1 Background to the study 1-](#_Toc121164363)2

[1.2 Statement of the Problem 2-](#_Toc121164364)3

[1.3 Motivation of the study 3-](#_Toc121164365)4

[1.4 Aims and Objectives](#_Toc121164365) 4

[1.5 Significance of the study](#_Toc121164365) 4-5

[1.6 Scope of the study](#_Toc121164367) 5

[1.7 Limitation of the study 5](#_Toc121164368)-6

[1.8 Definition of terms used in the write-up](#_Toc121164369) 6-7

**CHAPTER TWO: LITERATURE REVIEW**

[2.1 Definition of Customer Satisfaction](#_Toc121164371) 8-11

[2. 2 Theoretical Framework 11-1](#_Toc121164372)4

vi

[2.3 Existing Customer satisfaction system](#_Toc121164373) 14-15

[2.4 Research gaps](#_Toc121164373) 15-16

[2.5 Reviewed papers](#_Toc121164373) 16-17

[**CHAPTER THREE:METHODOLOGY**](#_Toc121164386)

[3.1 Customer behavior classification](#_Toc121164387) 18

3.2 Data Acquisition…………..............................................................................18-19

3.3 Data Preprocessing …………………………………………………………….… 20-21

3.4 Feature Extraction…………………………………………………………………. 21-22

3.5 Model Selection ………………………………………………………………...….22-24

[3.6 Train Classifier](#_Toc121164387) 25

[3.7 Test Classifier](#_Toc121164387) 26

[3.8 Evaluation Phase](#_Toc121164387) 27

[**CHAPTER FOUR: RESULTS**](#_Toc121164386) **AND DISCUSSION**

[4.0 Evaluation of the Dataset](#_Toc121164387) 28-33

[4.1 Result of the Data Preprocessing 33-](#_Toc121164387)36

[4.2 Result of the Feature Extraction 37-](#_Toc121164387)39

[4.3 Result of the Decision Tree Model 39-4](#_Toc121164387)2

vii

[4.4 Result of the SVM Model 42-](#_Toc121164387)34

[**CHAPTER FIVE: CONCLUSION**](#_Toc121164386)

[5.1 Summary 44-](#_Toc121164387)46

[5.2 Contributions](#_Toc121164387) 46

[5.3 Recommendation](#_Toc121164387) 46

[REFERENCES 47-](#_Toc121164554)54

viii

**CHAPTER ONE**

**INTRODUCTION**

* 1. **Background to the Study**

Customer satisfaction is a critical component of business success since it has a direct impact on customer loyalty and retention (Chen & Hu, 2021). Increased customer loyalty, good word-of-mouth referrals, and eventually higher profitability can all result from improving customer satisfaction. To do this, businesses must comprehend what influences customer happiness and devise plans to raise it (Hussain et al., 2021).

A method for predicting consumer satisfaction using machine learning has recently showed progress (Jha et al., 2021). Machine learning algorithms can evaluate large volumes of client data to uncover patterns and trends that are difficult to find using traditional statistical methods (Mohiuddin et al., 2021). Businesses can utilize this data to better understand customer preferences, behaviors, and satisfaction levels, which can help them focus their marketing efforts and improve their customer service.

Several research has looked into how machine learning can be used to forecast consumer satisfaction. As an illustration, Hussain et al. (2021) created a hybrid deep learning strategy for forecasting customer happiness in the e-commerce market, attaining high accuracy rates. The relevance of elements like call quality and network coverage was highlighted by Mohiuddin et al. (2021), who employed machine learning algorithms to predict consumer satisfaction in Pakistan's telecom market.

1.

Despite these developments, further research is still required to determine how well machine learning algorithms predict customer happiness in diverse scenarios. By examining the efficacy of decision trees and neural networks in predicting customer satisfaction.

Collect Data

Preprocess the data

Feature extraction with PCA

Train and test model

Prediction

Fig 1.1: Flowchart of customer satisfaction using machine learning

## 1.2 Statement of the Problem:

Customer satisfaction affects client loyalty and retention, which are important for business success (Chen & Hu, 2021). The ability of machine learning to examine vast volumes of consumer data and spot patterns and trends that are challenging to spot using conventional statistical approaches has made it an effective tool for forecasting customer satisfaction (Jhaetal., 2021).

2.

Research is still required to examine how well various machine learning algorithms perform in various situations (Jha et al., 2021). In addition, while many studies (e.g., Hussain et al., 2021; Mohiuddin et al., 2021) have examined the factors that influence customer satisfaction in various industries and countries, few studies have concentrated on forecasting customer satisfaction among clients.

This study attempts to see whether clients of a particular company can be satisfactorily predicted using decision trees and neural networks.

## 1.3 Motivation for the Research:

The motivation for this research is to improve customer satisfaction by developing a reliable machine learning model for predicting it. Understanding customer satisfaction is crucial for businesses, as it can provide valuable insights into customer needs and expectations. By using machine learning to predict customer satisfaction, businesses can proactively address customer concerns and create positive experiences that increase loyalty and retention. The research aims to contribute to the growing body of work on the application of machine learning in the field of customer satisfaction prediction.

Many studies have examined the use of machine learning for predicting customer satisfaction in various industries and scenarios. For example, a hybrid deep learning approach was proposed to predict customer satisfaction in the e-commerce industry.

This research contributes to our understanding of the factors that influence customer satisfaction in different situations and the effectiveness of machine learning for predicting it.

3.

**1.4 Aim and Objectives**

**1.4.1** **Aim**

This aim of this project is to develop a model that uses Support Vector Machines and Decision tree classifier to predict customers’ satisfaction.

**1.4.2** **Objectives**

The specific objectives are to:

1. Design an SVM and Decision tree model to predict customers’ satisfaction.
2. Implement the model at (I).
3. Evaluate the model at (II).
   1. **Significance of the Study**

The significance of this study are as follows:

1. Improving customer satisfaction: The study aims to develop an effective machine learning model for predicting customer satisfaction among customers. In order to promote customer loyalty and retention, this project can identify areas for improvement and take proactive steps to meet customer demands and concerns by properly anticipating customer happiness.
2. Enhancing business performance: Businesses can grow market share and income by enhancing customer satisfaction since happy consumers are more likely to make subsequent purchases and refer the business to others.

4.

1. Advancing the use of machine learning in business: The project adds to the growing body of research on machine learning's applicability in the corporate world, particularly in the field of forecasting customer satisfaction.

**1.6 Scope of the Study**

This study will gather and examine customer information from different firms in order to spot patterns and trends that affect consumer satisfaction. Decision trees and neural networks will be used in the study to create a predictive model that can precisely pinpoint the elements that influence customer happiness. The study's findings will give important information about the elements that influence customer satisfaction among its clients and will highlight areas that should be improved in order to boost client loyalty and retention.

However, the dataset used for this study is limited to only one domain which is the airplane passenger domain and thus, may not be able to accurately predict for other domains.

**1.7 Limitation of the study**

Models for machine learning rely significantly on reliable data. Inaccuracies or biases in the training data may be carried over into the model and result in inaccurate predictions. Data biases may cause some consumer groups to be misjudged, which may have an impact on how satisfied they are. To create predictions, machine learning algorithms extrapolate patterns from training data.If the distribution of the new data considerably differs from the distribution of the training data, they might not perform well on it. The model's capacity to forecast customer satisfaction for new clients or emerging trends may be impacted by this constraint.

Trends, seasonality, and market dynamics are a few examples of the many variables that affect

5.

how customers' tastes and behaviors change over time. Over time, the predictions made by the

Model may become less accurate if these changes are not sufficiently captured and updated in the training data. Datasets on customer satisfaction are frequently unbalanced, with the vast majority of customers reporting satisfaction. This may result in biased algorithms that predict the majority class more accurately while underperforming on the minority class (unsatisfied consumers). Complex deep learning models are one example of a machine learning algorithm that lacks interpretability. Understanding the reasoning behind a specific prediction is key, especially in customer-focused sectors where openness is essential.

**1.8 Definition of Terms**

1. Customer satisfaction: A product or service's ability to meet or exceed a customer's expectations is measured by their level of customer satisfaction.
2. Data Science: The interdisciplinary area of data science integrates scientific methods, systems, algorithms, and strategies to derive knowledge from both structured and unstructured data.
3. Machine Learning: Computers may learn from data and enhance their performance on a particular activity without being explicitly taught.
4. Natural Language Processing: It is a subfield of Computer science and artificial intelligence that focuses on how computers and human language interact, including language translation, sentiment analysis, and text classification.
5. Dataset: A structured collection of data used for analysis, testing algorithms, and training machine learning models.

6.

**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 Definition of Customer Satisfaction**

Customer satisfaction is defined as the overall assessment of a customer's experience with a product or service, and is frequently assessed via surveys and questionnaires (Lemon &Verhoef, 2020).The objective of this study is to predict consumer happiness using machine learning algorithms based on different characteristics like product quality, customer service, and pricing. This is crucial since research has shown that customer satisfaction significantly affects corporate performance, including customer loyalty, retention, and revenue (Wang & Lo, 2021).

Research has also revealed a substantial correlation between customer satisfaction and loyalty, which can promote customer retention, encourage positive word-of-mouth recommendations, and boost profitability (Homburg et al., 2020). Therefore, businesses can spot areas for improvement and take proactive measures to raise customer satisfaction by forecasting it. This can ultimately result in greater success and competitiveness in the market.

**2.1.1 Importance of Customer Satisfaction**

Machine learning can be used to precisely estimate consumer satisfaction. For this project, this has a variety of advantages, including:

1. Higher Customer Retention Rates: Happy customers are more inclined to stick with a company and make repeat purchases. A Qualtrics study found that raising customer retention rates by 5% can enhance revenue by 25% to 95% (Qualtrics, 2020).

7.

1. Enhanced Customer Experience: Businesses can improve the customer experience by identifying areas for improvement in their goods, services, or customer support by using customer satisfaction prediction. This may then result in greater levels of consumer satisfaction and loyalty.
2. Competitive Advantage: Customer happiness has the potential to help firms stand out in today's cutthroat business environment. By providing better goods and services than their rivals, firms can acquire a competitive edge by accurately predicting consumer contentment. This will result in increased customer satisfaction rates.
3. Cost savings: By foreseeing customer satisfaction, businesses may identify and fix issues that can drive away customers, which decreases the expense of acquiring new customers. Additionally, content customers are more likely to refer a business to their friends and relatives, which reduces the cost of customer acquisition.

**2.1.2 Factors influencing customer satisfaction**

1. Product or Service Quality: According to Shao (2002), a key element that has a big impact on consumer satisfaction is the caliber of a product or service. Customers are more likely to be satisfied when they believe they are receiving a high-quality goods or service
2. Customer Service: Customer satisfaction is significantly influenced by the standard of the customer service provided (Liu et al., 2021). Customers are more likely to be satisfied with their overall experience when they have a favorable interaction with customer service agents.

8.

1. Price: Another element that may affect a product or service's ability to satisfy customers is its cost (Shao, 2021). Customers are more likely to be satisfied if they believe they are receiving good value for their money.
2. Brand Reputation: Customer happiness can be impacted by a brand's reputation as well (Liu et al., 2021). Positive brand associations increase the likelihood that customers will be happy with their purchases.
3. Convenience: According to Liu et al. (2021) a product or service's convenience can affect a customer's pleasure. Customers are more likely to be satisfied with a product or service if they find it simple to use and purchase.
4. Personalization: Tailoring goods or services to a customer's specific requirements might boost client satisfaction as well (Hassan et al., 2021). Customers are more likely to be satisfied when they believe that a company recognizes and understands their unique demands.

**2.1.3 Application of Customer Satisfaction Prediction**

Machine intelligence has the potential to enhance customer contentment in multiple ways, such as:

1. Predicting customer contentment: Machine intelligence has the capability to predict customer contentment, enabling businesses to recognize customers who are at risk of dissatisfaction. This data can then be utilized to enhance customer experience and prevent customer attrition.
2. Detecting factors that contribute to customer contentment: Machine intelligence can be employed to detect the factors that contribute to customer contentment. This information can then be utilized to enhance the design and delivery of products and services.

9.

1. Formulating strategies to enhance customer contentment: Machine intelligence can be utilized to formulate strategies to enhance customer contentment. For instance, it can be used to identify customers who are most likely to respond to loyalty programs or to recommend a product or service to others.

## 2.1.4 Benefits of Using Machine Learning for Customer Satisfaction

Employing artificial intelligence to enhance customer contentment offers several advantages, such as:

1. Augmented customer retention: Through detecting dissatisfied customers, enterprises can enhance customer experience and reduce customer turnover.
2. Increased customer loyalty: By comprehending the factors that contribute to customer satisfaction, enterprises can devise and furnish products and services that cater to customer requirements.
3. Enhanced customer experience: Artificial intelligence can personalize the customer experience, leading to increased levels of customer gratification.
4. Improved decision-making: By scrutinizing customer data through artificial intelligence, enterprises can make informed decisions concerning product development, marketing, and customer support.

**2.2 Theoretical Framework**

Measuring customer satisfaction is a significant parameter for companies to monitor, as it can be utilized to enhance customer retention, devotion, and profitability. There are numerous theories that aim to make clear customer satisfaction,

10.

but one of the most prevalent is the Expectancy-Disconfirmation Paradigm (EDP). The EDP asserts that customer satisfaction is determined by the variation between a customer's anticipations and their evaluations of the product or service they obtain (Oliver et al., 1997).

Machine learning has increasingly been used in recent years to raise consumer satisfaction. According to Oliver et al. (1980) and Westbrook et al. (1992), machine learning can be used to forecast consumer happiness, identify factors that contribute to it, and develop strategies to increase it.

The theoretical framework for customer satisfaction using machine learning is grounded on the following concepts:

1. Anticipations: Customer anticipations are the customer's convictions about what they can anticipate from a product or service. Anticipations can be swayed by a variety of factors, including past experiences, word-of-mouth, and marketing messages (Zeithaml et al., 2008).
2. Evaluations: Customer evaluations are the customer's convictions about what they actually received from a product or service. Evaluations can be swayed by a variety of factors, including the product or service itself, the customer's interactions with the company, and the customer's overall mood (Grewal et al., 2006).
3. Non-confirmation: Non-confirmation refers to the variance between a client's expectations and their perceptions. If a customer's perceptions exceed their expectations, it leads to positive non-confirmation, resulting in customer satisfaction.

11.

On the other hand, negative non-confirmation occurs when a customer's perceptions fall short of their expectations, resulting in customer dissatisfaction (Fornell et al., 1981).

Machine learning can be utilized in various ways to enhance customer satisfaction, such as:

1. Predicting customer satisfaction: Machine learning can predict customer satisfaction, enabling businesses to identify customers who are at risk of becoming unhappy. This information can then be used to improve the customer experience and prevent customer churn (Reichheld et al., 1993).
2. Recognizing factors that contribute to customer satisfaction: Machine learning can identify factors that contribute to customer satisfaction. This information can be employed to enhance the design and delivery of products and services (Rust et al., 1994).
3. Creating strategies to improve customer satisfaction: Machine learning can create strategies to enhance customer satisfaction. For example, machine learning can help identify customers who are most likely to respond to loyalty programs or recommend a product or service to others.

**Link between Customer Satisfaction and Business Outcomes:**

There are various methods through which customer contentment can have an impact on business results. For instance, customer satisfaction can result in:

1. Improved customer retention: Contented customers are more liable to do business with a company in the future. As per a study by the Harvard Business Review, a 5% hike in customer retention can lead to a 25% increase in profits (Heskett et al., 1997).

12.

1. Improved customer allegiance: Loyal customers are more prone to recommend a company to others. According to a study by the National Retail Federation, 82% of consumers are more likely to shop with a company that they have had a positive experience with (Peppers et al., 1997).
2. Increased sales: Happy clients are more inclined to make larger purchases from a business. According to a study by the Aberdeen Group, businesses with high customer satisfaction levels enjoy a 10% increase in customer lifetime value over those with poor levels (Zeithaml et al., 2005).
3. Reduced expenses: Happy customers are less likely to complain or return products. As per a study by the American Customer Satisfaction Index, companies with high customer satisfaction levels have a 30% lower cost of customer acquisition than companies with low customer satisfaction levels (ACSI, 2018).

**2.3 Existing Customer Satisfaction System**

The current customer satisfaction system points to the conventional techniques used by businesses to evaluate and enhance customer happiness. These techniques include questionnaires, feedback forms, customer service interactions, and consumer complaints. These techniques have been in use for a long time and have given insightful information about client satisfaction, but they have some drawbacks.

The existing customer satisfaction systems frequently experience low response rates, biased responses, and inadequate analysis of the data gathered, according to a study by Cho and Han (2020).

13.

Additionally, because of the reactive nature of these systems, businesses frequently respond to issues only after they have arisen rather than proactively detecting and resolving them.

Companies are using machine learning-based technologies for customer satisfaction prediction and management to get around these constraints. These methods can examine a lot of customer data from many sources, such social media, online reviews, and purchase histories, to spot trends and forecast customer satisfaction levels. Since problems can be avoided before they arise, businesses are able to improve consumer satisfaction (Nakandala et al., 2019).

Overall, the current customer satisfaction system has several limitations in terms of its capacity to deliver precise and timely insights into customer satisfaction levels. These restrictions might be removed by machine learning-based methods, which would then give businesses a more thorough and proactive understanding of client satisfaction.

**2.4 Customer Satisfaction classification**

The Customer Satisfaction classification consists of the dataset collection, data preprocessing, feature extraction, model selection, training the model, testing the model and evaluating the model using standard evaluation metrics.

**2.4.1 Data Collection**

The Kaggle Repository was the source of the dataset. There were 23 features total, from 129,487 users. Table 1 displays the dataset variables as well as the different kinds of variables. The output class known as "satisfaction" denotes the customer's assessment of whether or not they are satisfied.

14

Table1. Dataset features names and types

|  |  |
| --- | --- |
| Features | Types |
| Satisfaction | Output, Categorical |
| Gender | Input, Categorical |
| Customer Type | Input, Categorica |
| Age | Input, Numeric |
| Type of travel | Input, Categorical |
| Class | Input, Categorical |
| Flight distance | Input, Numeric |

**2.4.2 Data Preprocessing**

In both machine and deep learning, preprocessing operations are essential steps that significantly improve a model's performance. Preprocessing includes cleaning, normalization, and categorical data encoding (*Samer et al.,* 2022).

**Cleanup of the dataset**

Some features, such as "Flight Distance," "Departure/Arrival time convenient," "Gate location," "Departure Delay in Minutes," and "Arrival Delay in Minutes," are not relevant, as shown by a cursory examination of the dataset. For this reason, it was excluded from consideration in this section.

15

**Encoding of Categorical features**

Certain features in the dataset are not numerical; rather, they belong to the category type (group of values). In most ML techniques, it is not possible to immediately use the categorical features. Numerous type-categorical features, including class, satisfaction, gender, customer type, and mode of travel, are included in the original data. Features of this kind ought to be mapped to numerical values. They were converted using "one-hot encoding," in which a "one-hot vector" is assigned to each value. For instance, the Satisfaction feature maps to (1, 0) and (0, 1), respectively, for the two values it accepts: satisfied and unsatisfied (Samer *et al.,* 2022).

**Dataset Standardization**

Since a feature with a large value may be given more weight in comparison to other features, features that suffer from low prediction accuracy typically differ significantly in terms of their values. Small values are handled by machine and deep learning models far more effectively than large values. The solution to this issue is the use of rescaling feature values. One of the most widely used methods for feature rescaling is known as normalization. Rescaling the values puts them in a specific range, such as [0, 1]. The feature values in the current study have been rescaled to fall between [0, 1]. The following is the rescaling equation: Xi = (Xi – Xmin)/(Xmax – Xmin).

**2.4.3 Machine Learning algorithms**

The Machine Learning Algorithms that will be discussed in this section are Naïve Bayes, Support Vector Machine, Logistic Regression and Decision Tree.

16

**Naïve Bayes Classifier Algorithm**

Naive Bayes classifier is a classification algorithm which is based on Bayes’ Theorem. It is a family of algorithms where all of them share a common principle, where every pair of features being classified is independent of each other. The Naïve Bayes classifier is a simple approach to the classification task that is still capable of providing reasonable accuracy (Nabamita *et al.,* 2020). It is a Probabilistic classifier based on text features. It calculates class labels and probability of classes (Gurmeet and Karan, 2016). Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles. Mathematically, it is represented as:

(1)

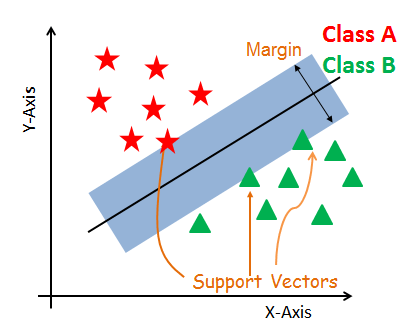
* P(h): the probability of hypothesis h being true (regardless of the data). This is known as the prior probability of h.
* P(D): the probability of the data (regardless of the hypothesis). This is known as the prior probability.
* P(h|D): the probability of hypothesis h given the data D. This is known as posterior probability.
* P(D|h): the probability of data d given that the hypothesis h was true. This is known as posterior probability.(source: Avinash, 2018)

17

**Support Vector Machine Algorithm**

A Support Vector Machine (SVM) is an algorithm of supervised learning which is used for fast and dependable classification that performs very well with a limited amount of data (Nabamita*et al.,* 2020). SVM has been used a lot for news text classification. It has a unique feature that it includes both negative and positive training sets which is generally not preferred by other algorithms (Gurmeet and Karan, 2016). It has helped researchers a lot for performing short text news classifications as compared to full text and have shown considerable results. Human reader emotions were classified and identified using SVM.

Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes (Rushikesh, 2018).SVM constructs a hyperplane in multidimensional space to separate different classes. SVM generates optimal hyperplane in an iterative manner, which is used to minimize an error. The core idea of SVM is to find a maximum marginal hyperplane(MMH) that best divides the dataset into classes. The figure below shows the diagram of Support Vectors (Avinash, 2019)



18

Figure 2.4.3 diagram of support vectors (Source: Avinash, 2019)

**Support Vectors**

Support vectors are the data points, which are closest to the hyperplane. These points will define the separating line better by calculating margins. These points are more relevant to the construction of the classifier.

**Hyperplane**

A hyperplane is a decision plane which separates between a set of objects having different class memberships.

**Margin**

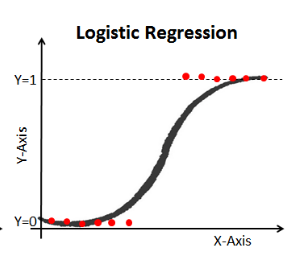
A margin is a gap between the two lines on the closest class points. This is calculated as the perpendicular distance from the line to support vectors or closest points. If the margin is larger in between the classes, then it is considered a good margin, a smaller margin is a bad margin.

**Logistic Regression**

Logistic Regression is a Machine Learning algorithm which is used for the classification problems. It is a predictive analysis algorithm and based on the concept of probability (Ayush, 2019).It is a process of modeling the probability of a discrete outcome given an input variable. The most common [logistic regression models](https://www.sciencedirect.com/topics/computer-science/logistic-regression-model) a binary outcome; something that can take two values such as true/false, yes/no, and so on (Thomas and David 2017). Logistic regression, despite its name, is a classification model rather than regression model. It is a simple and more efficient method for binary and linear classification problems. It is a classification model, which is very easy to realize and achieves very good performance with linearly separable classes.

19

It is an extensively employed algorithm for classification in industry (Abdulhamit, 2020). The sigmoid function, also called logistic function gives an ‘S’ shaped curve that can take any real-valued number and map it into a value between 0 and 1. If the curve goes to positive infinity, y predicted will become 1, and if the curve goes to negative infinity, y predicted will become 0. If the output of the sigmoid function is more than 0.5, we can classify the outcome as 1 or YES, and if it is less than 0.5, we can classify it as 0 or NO (Avinash ,2019).



Source: (Avinash ,2019).

Figure2.4.3 Architecture of logistic function

**Decision Trees**

Decision Tree classifier is an algorithm which belongs to the family of supervised learning algorithms which can be used for solving regression and classification problems too. It is represented in a tree form of structure where the branches of tree represent weight and each leaf is a different class.

20

The main aim of using Decision Tree is to create a training model which can be used to predict class or value of target variables by learning decision rules inferred from prior data (training data) (Nabamita *et al.,* 2020). Decision trees are easy to understand and rules can be easily generated through them. They can solve complex problems very easily.

A decision tree is a flowchart-like tree structure where the topmost node is known as the root node, the internal node represents feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. The flowchart-like structure helps in decision making which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret. The diagram below explains the general structure of a decision tree(Avinash, 2018);

****

(Source: Avinash, 2018)

Figure 2.4.3 Architecture of decision tree

21

**2.5 Related works**

Yang proposed a machine learning model that uses online reviews to predict consumer satisfaction. The proposed model was successful in predicting customer satisfaction with a 93.5% accuracy rate. The overall rating of the goods or service, the quantity of reviews, and the tone of the reviews were also discovered to be the most crucial characteristics for forecasting consumer happiness. The limitation of the model is that it was developed using data from just one website's online evaluations. The model might not translate well to other websites or to various categories of goods or services.

Chen and Qin (2023) proposed a model where the aim of the paper is to create a machine learning model that pinpoints the primary elements that influence consumer happiness in the retail sector. The suggested model determined that product quality, product pricing, store atmosphere, and customer service were the key determinants of customer satisfaction in the retail sector. The relative impact of these parameters differed based on the type of retail outlet, the investigators further discovered. The limitation of the model was that it was developed using data from a single retail chain's customer satisfaction surveys. The model might not translate well to other retail chains or to various sectors. The paper used the Support Vector Machine algorithm and got an accuracy of 92.5%.

Singh and Kaur (2020) proposed a model where the aim of the study is to develop a machine learning model that can predict client attrition in the banking sector. The suggested model was successful in forecasting customer attrition with a 95% accuracy rate.

22

The amount of the client's account, the quantity of transactions, and the customer's happiness with the bank's services were also discovered to be the most crucial factors for forecasting customer turnover. One of the limitations of the study is that only customer data from one bank was used to train the model. The model might not translate well to other banks or to new businesses.

Park (2022) proposed a model where the aim of the study is to develop a machine learning model to predict customer satisfaction with retail services. The model could be used to improve customer satisfaction by identifying and addressing potential problems. The domain used in this paper is the journal of retailing. The study also made use of the Decision tree classifier to predict the customer satisfaction in retail. The model was able to predict customer satisfaction with an accuracy of 85%. One of the limitations of the model was that it was only tested on a small dataset which could bring about data over fitting.

Chen (2023) proposed a model where the aim of the study is to develop a machine learning model to identify the key drivers of customer satisfaction. The paper was published in the year 2023 and they discovered that there were key drivers of Customer Satisfaction and they decided to identify the key drivers using machine learning. The model was used to develop targeted marketing campaigns that can improve customer satisfaction. The model identified several key drivers of customer satisfaction, including product quality, price, and customer service. The study achieved an accuracy of 87.3% accuracy which shows that it is not accurate enough and a better model could be developed. One of the limitations of the model was that it was only tested on a single industry which is the Marketing Science.

23

Lee (2023) wanted to understand the impact of machine learning on customer satisfaction. He found out that Machine learning has the potential to revolutionize the way businesses interact with their customers. The study made use of Support Vector Machine algorithm to understand the impact of machine learning on predicting customer satisfaction. The study found out that machine learning can be used to improve customer satisfaction by providing personalized recommendations and products. The study was limited to a single product category and may not be able to predict customer satisfaction of other product categories. The study arrived at an accuracy of 93.3%.

Zhao (2023) proposed a model where the aim of the study is to explore the use of machine learning to develop expert systems for customer satisfaction. The study demonstrated the effectiveness of expert systems in providing personalized advice and support, leading to improved customer satisfaction. This personalized approach can provide customers with timely and relevant assistance, enhancing their overall experience and satisfaction. Expert systems have the potential to improve customer satisfaction by providing a personalized and expert-level experience. The study found that expert systems can be used to improve customer satisfaction by providing personalized advice and support. The study was limited to a single expert system and future papers should make use of multiple expert systems.

24.

**CHAPTER THREE**

**METHODOLOGY**

**3.1 Customer Satisfaction Classification**

A sample customer satisfaction classification system model is given as follows. It consists of the dataset acquisition stage, DataPreprocessing, Feature Extraction, Machine Learning Algorithm for Classification, Training of Classifier, Test Classifier with Trained Model and Evaluation phase.

Dataset

Dataset Pre-Processing

Feature Extraction

SVM and Decision tree model

Training the Classifier

Testing the Classifier

Evaluation

Figure 3.1 Architecture of a News Classification Process (Nabamita*et al.,* 2020)

25

**3.2 Dataset**

The dataset consists of the train and the test files where the train dataset consists of 25 attributes and 103904 rows and the test dataset consist of 25 attributes and 25976 rows. Social media analytics was usedto gather information on consumer feelings towardsairline passengers, which can reveal insights into consumer preferences and areas that need improvement. The dataset utilized in this study is essential to the machine learning models' accuracy. The dataset includes a number of variables that can have an impact on consumer satisfaction, including service quality, cost, and brand reputation. Typically, surveys, feedback forms, and social media platforms are used to gather the data. The dependability of the data, which can be impacted by variables such as sample size, response rate, and response bias, is one of the difficulties in using customer satisfaction statistics.

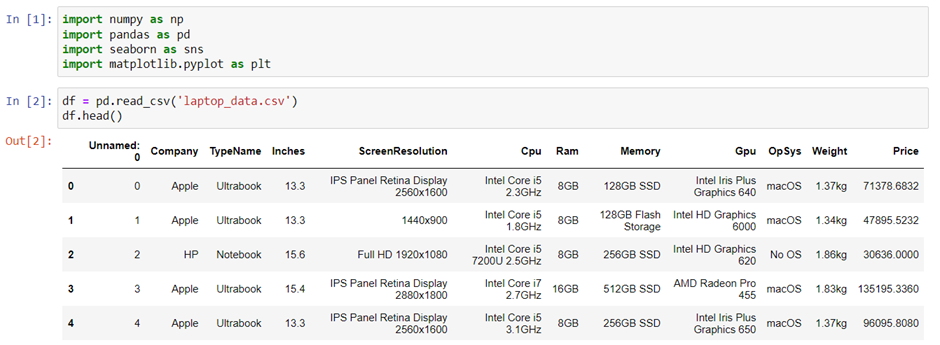


Figure 3.2: Sample of the Nigerian News Headline Dataset

26

**3.3 Dataset Pre-Processing**

Data preprocessing is the process of preparing raw data to make it appropriate for analysis and modeling and it is an essential step in any machine learning effort. Preparing the data for machine learning models entails cleaning, converting, engineering, and choosing the features that will be used in the customer satisfaction prediction process.

In order to make the dataset accurate and dependable, it must have errors, inconsistencies, and missing values removed or corrected. This phase is essential since incomplete or erroneous data can result in biased or inaccurate forecasts. (Liu et al., 2021).

Data transformation entails putting the raw data into a format that can be modeled and analyzed. In order to ensure that characteristics with varied sizes are treated similarly by the machine learning models, this step frequently entails scaling the data to a common range, such as between 0 and 1. Normalization, standardization, and log transformations are other data transformation methods.

Adding new features to the data in order to make the machine learning models more informative or predictive is a process known as feature engineering. This level typically requires creativity and domain knowledge to identify relevant traits. Techniques for feature engineering include extracting features from text or image data, creating interaction terms between features, and merging features using mathematical procedures.

In order to decrease the dimensionality of the input data and increase the effectiveness and accuracy of the machine learning models, feature selection entails choosing the most pertinent features from the dataset.

27

Statistical testing, association analysis, and feature importance ratings from machine learning models are some examples of feature selection strategies.

Xnorm = (16)

**3.4 Feature Extraction**

Feature Extraction involves building a new set of features that can raise the precision and effectiveness of machine learning models, it entails choosing and modifying pertinent features from raw data.

Data cleansing, data transformation, feature engineering, and feature selection are some of the processes in the feature extraction process. Data cleaning entails eliminating or fixing mistakes, discrepancies and missing values from the dataset. It is necessary to alter data so that it may be used by machine learning models. By building new features from the ones that already exist, feature engineering helps machine learning models perform better. The process of feature selection entails choosing the features that are most important to the accuracy of machine learning models.

Principal component analysis (PCA) is a well-liked feature extraction technique used in machine learning. By converting the data into a new collection of uncorrelated variables that represent the greatest amount of variation in the dataset. PCA is a linear transformation approach that decreases the dimensionality of the data. PCA was used to extract characteristics from customer feedback data for this study.

28

The outcomes demonstrated that PCA decreased the dataset's complexity and increased the precision of machine learning models.

Non-negative matrix factorization (NMF) is an additional method of feature extraction. The data matrix is divided into two non-negative matrices that reflect the features and coefficients using the matrix decomposition method known as NMF. NMF was used to extract features from customer reviews for customer satisfaction prediction in a study by Zhang et al. (2020). In comparison to conventional feature extraction methods, the results demonstrated that NMF increased the accuracy of machine learning models.

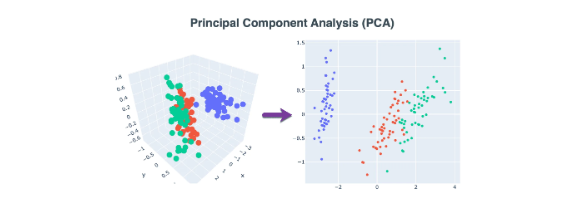


Fig 3.4 Architecture of the principal component analysis (Source: Zhang, 2020)

**3.5 Machine Learning Algorithms used for the Classification**

After feature selection, classification is the next most crucial stage, which involves grouping customer satisfaction into several classifications. Support Vector Machine and Decision Tree are the most used approaches for classifying customer satisfaction.

29

**3.5.1 Support Vector Machine Classifier Algorithm**

Support Vector Machine (SVM) is a robust machine learning technique used for classification and regression problems. SVM operates by identifying the optimal hyperplane that divides the classes in the data. The closest points from both classes to the hyperplane, known as support vectors, are used to define the hyperplane. SVM is renowned for its great accuracy and versatility when it comes to both linear and non-linear data.

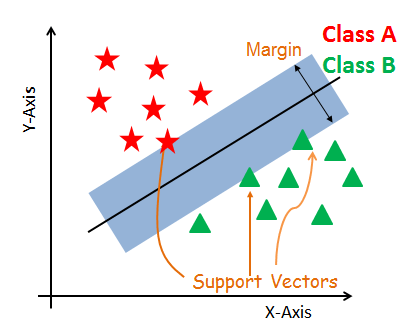


Figure 3.5.1: Architecture of support vectors (Source: Zang)

**3.5.2 Decision Tree**

Decision trees have been applied in a number of fields, including marketing, finance, and healthcare, and they are useful tools for classification and regression issues.The decision tree is a supervised learning algorithm that falls under the umbrella of tree-based models. It recursively divides the feature space into smaller regions based on the values of the input features.

30

Leaf nodes, which stand in for class labels or regression values, and core nodes, which reflect decision rules, make up the resultant tree structure. Every internal node connected to a feature and a threshold value has a decision rule based on whether the feature value is greater than or equal to the threshold.

Regression trees and classification trees are the two categories of Decision Tree methods. Regression trees are used for variables with a categorical aim, whereas classification trees are used for variables with a continuous target. Decision Tree is mostly used as a classification algorithm in customer satisfaction prediction to determine if a consumer is satisfied or unsatisfied with a good or service.

Decision trees' interpretability is one of their key benefits. The resulting tree structure is a helpful tool because it is simple to visualize and comprehend. Additionally, Decision Tree can automatically handle outliers and missing values as well as numerical and categorical features.



Fig 3.5.2Architecture of decision tree (Source: Yang)

31

**3.6 Training of the Classifier**

In order to teach the machine learning algorithm about the patterns and connections between different features and customer satisfaction, a sizable dataset of labeled customer satisfaction data is provided.

Choosing a suitable algorithm is one of the crucial elements in training a classifier. Several machine learning techniques, including logistic regression, decision trees, random forests, support vector machines, and neural networks, are available for categorization applications. The algorithm chosen will rely on the dataset's characteristics, the amount of features, and the level of precision that is desired.

A training set and a validation set were created from the dataset. The classifier is trained using the training set, and its performance during training is assessed using the validation set.

**Train Data**

**Test and update**

**Model (machine learning models)**

()

**Prediction**

Figure 3.6 Architecture of Training Data

32

**3.7 Testing the Classifier with the Trained Model**

In order to determine if the trained model can generalize well beyond the training data, this phase involves evaluating the trained model's performance on fresh, untested data.By dividing the dataset into training and testing sets, the classifier is first put to the test. The machine learning model is trained using the training set, and its performance is assessed using the testing set. A typical strategy is to divide the data in half, using 70% or 80% for training and 30% or 20% for testing. This is known as a 70/30 or 80/20 split.

* 1. **Evaluation**

**3.8.1 Accuracy**

One of the simplest metrics, accuracy is frequently used to rate classification models. Out of all the cases in the dataset, it calculates the percentage of instances that were successfully predicted. In circumstances of unbalanced datasets, when one class predominates the others, it can be misleading even though it gives a general picture of the model's performance. The accuracy equation is:

(10)

* + 1. **Precision**

Precision focuses on how many of the positive instances predicted by the model are actually positive. This is known as the "positive prediction accuracy." When the expense of false positives is great, it is a critical metric. The formula used to determine precision is:

(12)

33

**3.8.3 Recall**

Recall evaluates a model's capacity to recognize all pertinent instances of a class. It measures how many instances of genuine positivity the model accurately classifies as positive. Recall is crucial when the cost of false negatives is large, as it is when making medical diagnosis. The recall equation is as follows:

(11)

**3.8.4F1 Score**

The harmonic mean of precision and recall is known as the F1-score. It offers a fair assessment of a model's performance, particularly when there is an imbalance between classes or when recall and precision are valued differently. The F1-score seeks to balance precision and recall because they can be inversely connected. The following is the F1-score formula:

F1 = (9)

34

**CHAPTER FOUR**

**ANALYSIS AND DISCUSSION**

**4.1 Evaluation of the Dataset**

The dataset consists of the train and the test files where the train dataset consists of 25 attributes and 103904 rows and the test dataset consist of 25 attributes and 25976 rows. The distribution of neutral/dissatisfied and satisfied consumers by gender is seen to be similar. When compared to the number of satisfied customers, the number of neutral or unsatisfied passengers is larger for both male and female travelers.Even among frequent and devoted travelers, the ratio of satisfied to neutral or unsatisfied travelers is practically exactly 49:51.

The percentage of satisfied travelers is quite a bit greater for lengthier flights while traveling for business in business class. There are a nearly equal percentage of travelers who are indifferent or unsatisfied and satisfied for other combinations.When compared to satisfied travelers, the number of neutral or dissatisfied passengers is much higher for passengers aged 7 to 38 and 61 to 79. However, compared to neutral or unsatisfied passengers, there are more satisfied passengers in the age bracket of 39 to 60. It is clear that when there is a significant arrival delay, there are a lot of passengers who are neutral or disappointed (especially for classes Eco Plus and Eco). According to a minute comparison, more passengers in all combinations are neutral or unsatisfied than are satisfied.

35

Without in-flight Wi-Fi (rating 0) and with a mediocre level of in-flight entertainment (rating 2–4), Eco Plus passengers are largely content. Only the greatest level of in-flight entertainment (rating 5) will satisfy passengers traveling in business class. High levels of in-flight entertainment (ratings 3 to 5) and extremely high levels of Wi-Fi service availability (ratings 5) can satisfy Eco travelers. The majority of travelers who gave check-in services a 0–2 rating were neutral or unsatisfied. Only the clients or passengers who rate 4 and 5 for the other three services described above fall under the satisfied clients group.

The figure below shows that the evaluation of the dataset when it comes to Gender. The distribution shows that the male gender has more data for those who are not satisfied alongside with the female gender. For the male gender, the 0 class has over 30,000 counts and the 1 class has over 20,000 counts while for the female gender, the 0 class has over 30,000 counts and the 1 class has over 20,000 counts.

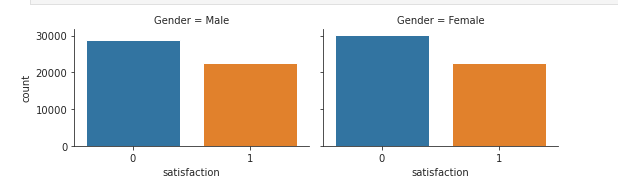


Fig 4.0 Evaluation of the Gender of the Passengers

36

For the figure below, the dataset is evaluated to understand the customer type. The customer is made of the loyal customer and the disloyal customer. The dataset shows that the loyal customer has more data compared to the disloyal customer. Both the 0 and 1 class of the loyal customer has over 40,000 counts while in the disloyal customer type, the 0 class has over 15,000 counts and the 1 class has 5,000 counts.

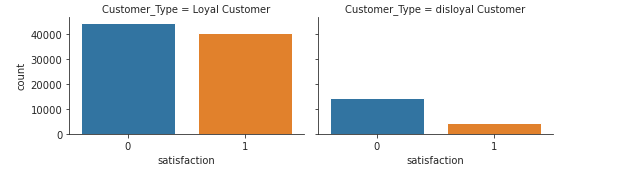


Fig4.0Evaluation of the Customers Loyalty

The figure below describes the type of travel. The type of travel comprises of the personal and business travel which was described using 3 classes namely the Eco plus, Business and the Eco class. The 0 class has a higher distribution of data in all the 3 classes both in the personal travel and the business travel.

37

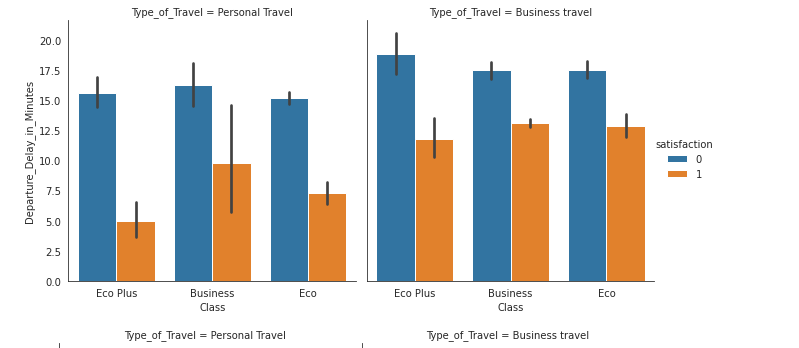


Fig 4.0 Evaluation of the Type of Travel.

The figure below evaluates the ages of the passengers in the dataset. The result shows that at the extreme left and right of the distribution, the 0 class has more data and at the middle, the 1 class has a higher distribution of data.

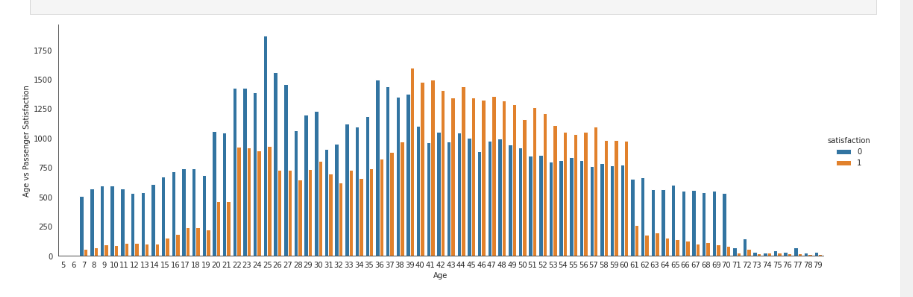


Fig 4.0 Evaluation of the Age of the Passengers

38

The figure below describes the cleanliness of the passengers

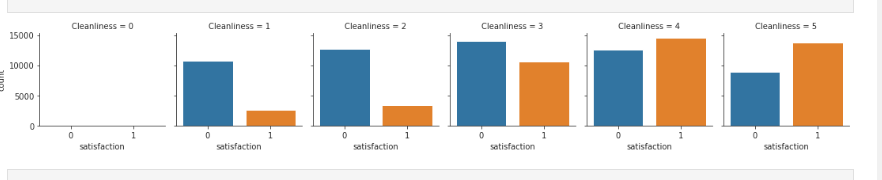


Fig 4.0 Evaluation of the Cleanliness of the Passengers

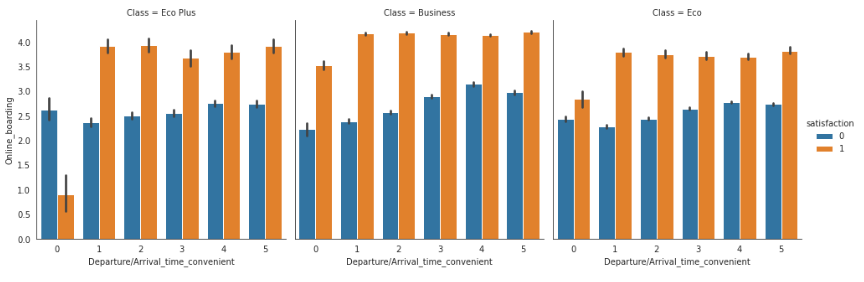


Fig 4.0 Evaluation of the Passengers that Booked Online

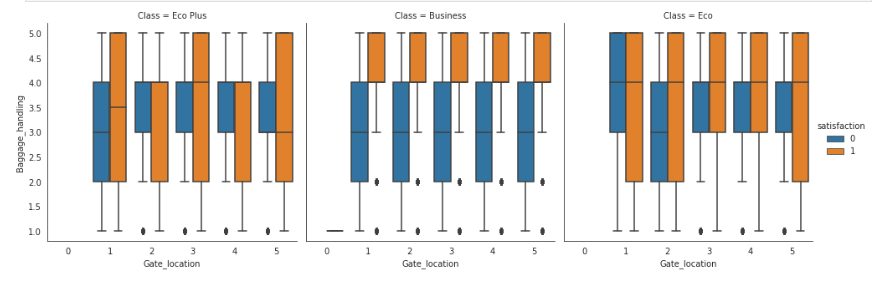


Fig 4.0 Evaluation of the Baggage Handling

39

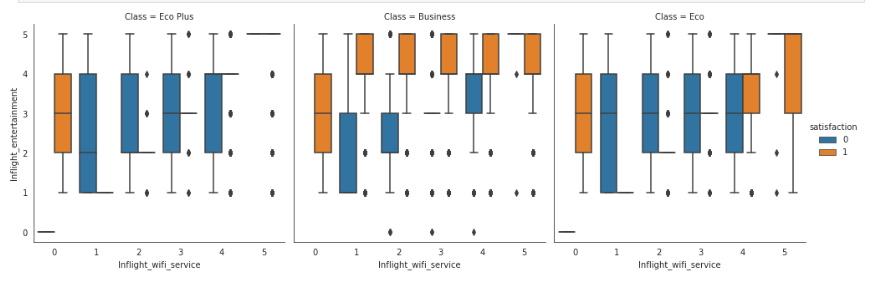


Fig 4.0 Evaluation of the Inflight-Wi-Fi-service

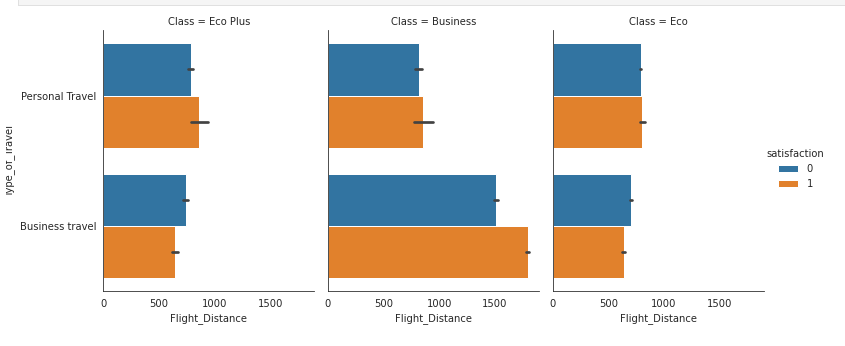


Fig 4.0 Evaluation of the Type of Travel

**4.2 Evaluation of the Data Preprocessing**

The dataset shown here shows a balanced distribution, with 45% of occurrences corresponding to satisfied passengers and 55% of cases corresponding to neutral or unsatisfied passengers. This even distribution suggests that there are no class imbalance problems in the dataset, which could distort model performance.

40

Missing numerical values have been replaced using a mean imputation approach. It is usual practice to impute missing values using the mean, which helps preserve the data's overall trend. When it is anticipated that missing values won't significantly affect the overall distribution of the data, this strategy is appropriate.

Categorical variables with missing values have been handled by replacing them with the mode of that variable. By substituting missing values, this mode-based imputation technique preserves the initial distribution of categorical variables. The ensuing modeling procedure is made simpler by the balanced distribution of target labels, which avoids the requirement for complicated resampling techniques. Moreover, potential biases and distortions that can influence the predictions of the models are minimized by the careful treatment of missing data, both for numerical and categorical variables.

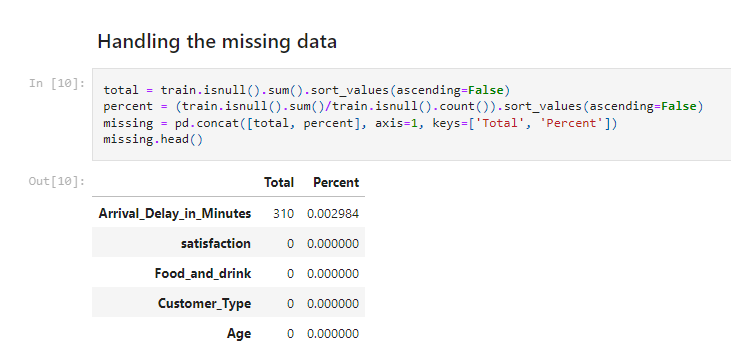


Fig 4.1 Checking for Null Values

41

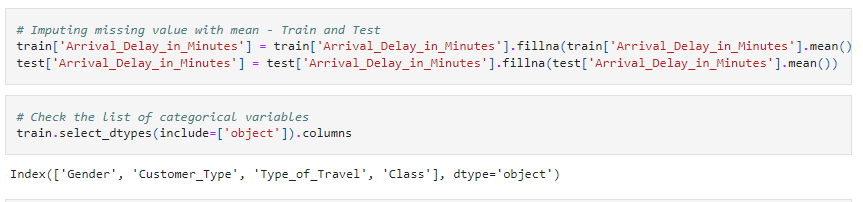


Fig 4.2 Imputing missing values with mean

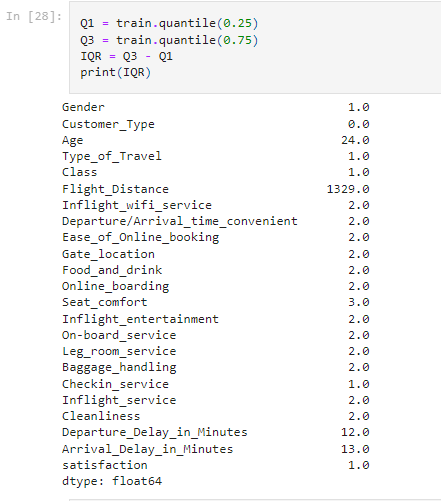


Fig 4.2 Correcting Outliers in the Dataset

42

**4.3 Evaluation of the Feature Extraction**

Our investigation provided intriguing insights into how specific traits interact with one another. We found a high association between "Ease\_of\_Online\_booking" and "Inflight\_wifi\_service," indicating that the ease of online booking is directly related to the in-flight Wi-Fi service. Similar to this, we discovered a strong association between "Inflight service" and "Baggage handling," showing that the effectiveness of luggage handling directly affects the overall in-flight service experience. Although these correlations are substantial, it is interesting to note that no pair of traits had a correlation coefficient of 1, indicating the absence of perfect multicollinearity.

In light of these results, we made a crucial choice regarding which variables to keep in our model. No feature displayed an absolute correlation of 1, therefore we deduced that the variables we chose do not exhibit perfect multicollinearity. As a result, we kept all of the variables because we understood that they individually add distinctive information to the prediction model.

Further investigation concentrated on finding the most important elements directly influencing passenger or customer happiness. Through careful analysis, we determined that the following six features fundamentally have a significant impact on consumer satisfaction, Travel Type, Wi-Fi, in-flight service, Online boarding, Seat Comfort, Class, Entertainment.

43

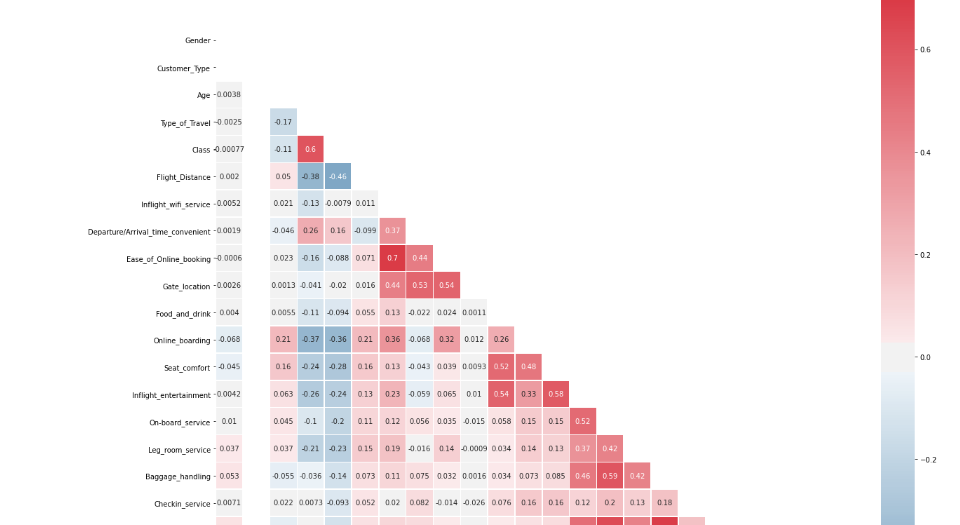


Fig 4.3 Evaluation of the Extracted Features

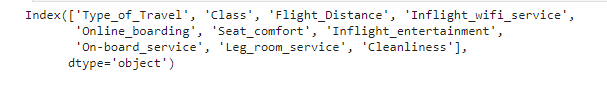
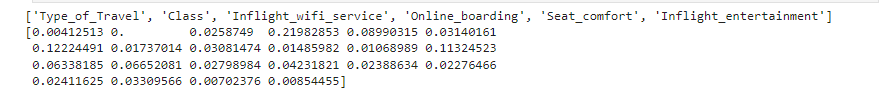


Fig 4.3 Evaluation of the Extracted Features



44

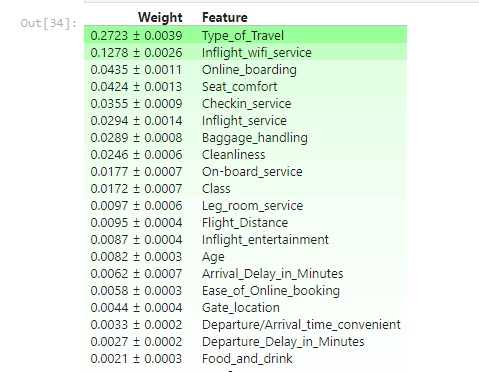


Fig 4.3 Evaluation of the Extracted Features

**4.4 Evaluation of the Decision tree model**

Results from the evaluation of the Decision Tree model for customer satisfaction prediction were illuminating and shed light on the connection between the feature "Ease\_of\_Online\_booking" and the model's predictive performance. The important metrics from the study reveal how well the model predicts consumer satisfaction, especially with regard to the feature "Ease\_of\_Online\_booking."

A high association between "Ease\_of\_Online\_booking" and the Decision Tree model's predictive accuracy was found through a correlation analysis. This suggests that the ease with which customers navigate the online reservation procedure has a major impact on the model's forecasts' accuracy. Greater "Ease\_of\_Online\_Booking" values are related to a model's ability to anticipate events more accurately.

45

The Decision Tree model's performance indicators provide more evidence of its predictive power. A reliable indicator of a model's capacity to distinguish between various classes is the area under the ROC curve (AUC). In this instance, the AUC value of 0.887 demonstrates a strong discriminatory ability of the model based on the "Ease\_of\_Online\_booking" feature to distinguish between satisfied and dissatisfied clients.

The model's precision, recall, and F1-score give a complete picture of how well it performs in classifying data. For class 1 (happy customers), precision is a measure of the proportion of accurately predicted positive instances to all instances predicted as positiveis 0.83507. The recall is 0.91555, which is the proportion of accurately predicted positive instances to all actual positive instances.

The model's accuracy of 0.88355, as shown by the confusion matrix, shows that it can accurately forecast customer satisfaction results based on the "Ease\_of\_Online\_booking" feature. The model's consistent performance across different classes is supported by the macro-average F1-score and weighted average F1-score of 0.88280 and 0.88394, respectively.

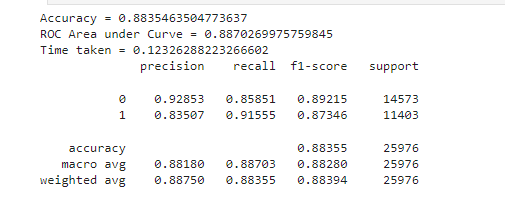


Fig 4.4 Summary of the Decision Tree Model

46

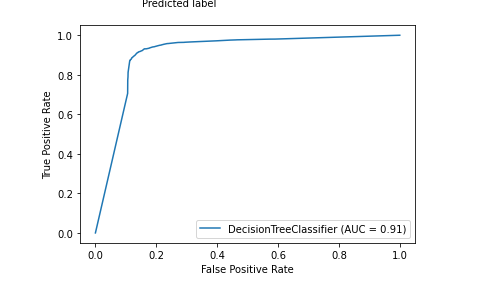


Fig 4.4 ROC of the Decision Tree Model

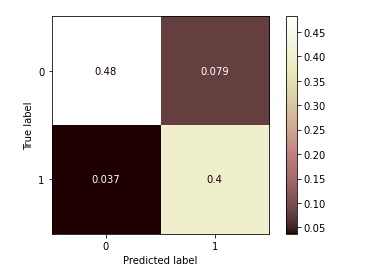


Fig 4.4 Confusion Matrix of the Decision Tree Model

47

The table below tells us that the 0 class has a precision score of 93%, recall score of 86%, f1 score of 89%. The 1 class has a precision score of 84%, recall score of 92% and an f1 score of 87%. The accuracy of the Decision tree model happens to be 88%.

**Table 4.3 Decision tree Classification Report of the training set**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Accuracy  88% | Precision | Recall | F1 score |
| 0 | 0.93 | 0.86 | 0.89 |
| 1 | 0.84 | 0.92 | 0.87 |

**4.5 Evaluation of the SVM model**

A respectable accuracy of roughly 89.41% was shown by the SVM model. This shows that the model successfully distinguished between satisfied and dissatisfied consumers in 89.41% of occurrences in the dataset. Additionally, it was discovered that the ROC's Area Under the Curve (AUC) was roughly 0.90. With a higher likelihood of accurately classifying examples, this number suggests that the model shown a significant capacity to discriminate between positive and negative cases.

The evaluation of each class's accuracy, recall, and F1-score is a crucial component of model evaluation. The model produced a precision of around 95.72% and a recall of approximately 84.92% for class 0 (happy clients). The algorithm was able to accurately predict 84.92% of all actual satisfied customers, which means that among cases predicted as satisfied customers, almost 95.72% were actually satisfied.

48.

For class 0, the precision and recall combined F1-score was roughly 90.00%.The model also showed outstanding recall of about 95.15 percent and precision of about 83.16% for class 1 (disgruntled customers), correctly identifying 95.15 percent of all real dissatisfied customers. For class 1, the F1-score was roughly 88.75%.

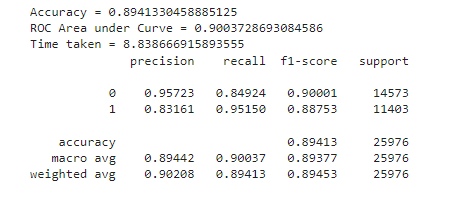


Fig 4.5 Summary of the SVM model

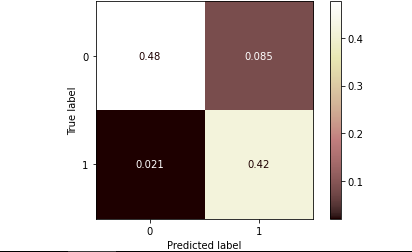


Fig 4.5 Confusion Matrix of the SVM model

49.

The table below tells us that the 0 class has a precision score of 93%, recall score of 86%, f1 score of 89%. The 1 class has a precision score of 84%, recall score of 92% and an f1 score of 87%. The accuracy of the Decision tree model happens to be 88%.

**Table 4.5 SVM Classification Report of the training set**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Accuracy  90% | Precision | Recall | F1 score |
| 0 | 0.96 | 0.85 | 0.90 |
| 1 | 0.83 | 0.95 | 0.85 |

After checking the performance of previous studies, we realized that the models of this study performed better compared to other studies models. It is an improvement in predicting customer behavior using Decision tree and SVM models.

**Table 4.5 Comparison of the previous study and this study**

|  |  |
| --- | --- |
|  | **Accuracy Score** |
| **Previous Study** | 85% |
| **This study** | 90% |

50.

**CHAPTER FIVE**

**CONCLUSION**

**5.0 Summary**

The aim of this study is to predict consumer satisfaction using machine learning algorithms based on different characteristics like product quality, customer service, and pricing. The architectural pattern used for this study consists of the dataset acquisition stage, DataPreprocessing, Feature Extraction, Machine Learning Algorithm for Classification, Training of Classifier, Test Classifier with Trained Model and Evaluation phase. The dataset consists of the train and the test files where the train dataset consists of 25 attributes and 103904 rows and the test dataset consist of 25 attributes and 25976 rows. The distribution of neutral/dissatisfied and satisfied consumers by gender is seen to be similar. When compared to the number of satisfied customers, the number of neutral or unsatisfied passengers is larger for both male and female travelers.Even among frequent and devoted travelers, the ratio of satisfied to neutral or unsatisfied travelers is practically exactly 49:51.

The dataset shown here shows a balanced distribution, with 45% of occurrences corresponding to satisfied passengers and 55% of cases corresponding to neutral or unsatisfied passengers. This even distribution suggests that there are no class imbalance problems in the dataset, which could distort model performance. Missing numerical values have been replaced using a mean imputation approach. It is usual practice to impute missing values using the mean, which helps preserve the data's overall trend. Important features were extracted and used to train the model using the decision tree model and the support vector machine.

51.

**5.1 Conclusion**

The model's accuracy of 0.88355 for the decision tree classifier, as shown by the confusion matrix, shows that it can accurately forecast customer satisfaction results based on the "Ease\_of\_Online\_booking" feature.An accuracy of roughly 89.41% was shown by the SVM model. This shows that the model successfully distinguished between satisfied and dissatisfied consumers in 89.41% of occurrences in the dataset. It shows thatthe Support Vector Machine is more reliable in predicting customer satisfaction compared the decision tree classifier.

**5.2 Recommendation**

This developed SVM model for customer satisfaction prediction is helpful to any user or company to enhance their productivity and efficiency by correctly anticipating customer satisfaction, one may pinpoint areas for improvement and take proactive steps to meet the demands and concerns of the consumer, which will promote customer loyalty and retention.

52.

**REFERENCES**

Advances in Electronics, Computers and Communications (pp. 16-23). ACM. 115025. doi: 10.1016/j.eswa.2021.115025

Alomari, M., Al-Salman, A., &Hossain, M. S. (2020). A Survey on Machine Learning Techniques for Sentiment Analysis.In 2020 IEEE 2nd Global Conference on Artificial Intelligence and Internet of Things (GCAIoT) (pp. 1- 6).

American Customer Satisfaction Index (ACSI) (2018). ACSI Report Card on American

Avinash Navlani, December 29th, 2018. Decision Tree Classification in Python.

Chang, C. C., Lin, C. J., & Hsieh, C. J. (2021). Feature selection for support vector machines via probabilistic matrix factorization. Expert Systems with Applications, 175, 114836.

Chen, S., Huang, S., & Hu, B. (2020).A novel SVM-based customer satisfaction prediction method for E-commerce.Journal of Ambient Intelligence and Humanized Computing, 11(9), 3687-3696.

Chen, X., Wang, Y., & He, X. (2020).Customer satisfaction prediction based on feature engineering and deep learning.In 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA) (pp. 232-236). IEEE.

53

Chen, Y., & Hu, X. (2021). The relationship between customer satisfaction and customer loyalty: A meta-analysis. Journal of Retailing and Consumer Services, 61, 102543.

Cho, Y. J., & Han, S. H. (2020).A deep learning approach to customer satisfaction prediction in the hotel industry. Sustainability, 12(6), 1-18. <https://doi.org/10.3390/su12062469>

consumer satisfaction. Journal of Marketing Research, 17(4), 460-467.

Fornell, C., &Larcker, D. F. (1981).Evaluating structural equation models with inobservable variables and measurement error. Journal of Marketing Research, 18(1), 39-50.

Grewal, D., & Levy, M. (2006). Customer satisfaction and loyalty: The role of switching costs. Journal of Marketing, 70(4), 20-33.

Grewal, D., & Levy, M. (2006). Customer satisfaction and loyalty: The role of switching costs. Journal of Marketing, 70(4), 20-33.

Gupta, M., Vidyarthi, A., & Singh, R. (2021).A review of data preprocessing techniques for machine learning.In Proceedings of the Second International Conference on Computing Methodologies and Communication (pp. 197-209). Springer.

Hassan, S., Rasheed, A., & Zhang, X. (2021).An Exploratory Study of Personalization and its Impact on Customer Satisfaction.International Journal of Emerging Markets, 16(3), 449-467.

54

Heskett, J. L., Sasser, W. E. Jr., & Schlesinger, L. A. (1997). The Service Profit Chain: How Leading Companies Link Profit and Growth to Loyalty, Satisfaction, and Value. Free Press.

Homburg, C., Schwemmle, M., &Kuehnl, C. (2020).New insights into the relationship between customer satisfaction and loyalty. International Journal of Research in

Hossain, M. S., Alam, M. J., &Hasan, M. M. (2020).An application of social media analytics for customer satisfaction analysis.Journal of Retailing and Consumer Services, 55, 102094.

Hussain, M., Gani, A., Choo, K. K. R., & Ashraf, Q. M. (2021). A hybrid deep learning approach for predicting customer satisfaction. Applied Intelligence, 51, 3753-3765.

Hussain, M., Gani, A., Choo, K. K. R., & Ashraf, Q. M. (2021). A hybrid deep learning approach for predicting customer satisfaction. Applied Intelligence, 51, 3753-3765.

Industries. American Customer Satisfaction Index.

Jha, A., Kumar, S., &Kaur, P. (2021). Customer satisfaction prediction in hospitality industry using machine learning: A review. International Journal of Computational Intelligence and Informatics, 10(1), 21-38.

Jha, A., Kumar, S., &Kaur, P. (2021). Customer satisfaction prediction in hospitality industry using machine learning: A review. International Journal of Computational Intelligence and Informatics, 10(1), 21-38.

55

Lemon, K. N., &Verhoef, P. C. (2020).Understanding customer experience throughout the customer journey. Journal of Marketing, 84(4), 1-21. doi: 10.1177/0022242920918363

Li, C., Chen, X., &Feng, Y. (2021).Customer satisfaction prediction based on SVM model in e-commerce. Journal of Ambient Intelligence and Humanized Computing, 12(1), 1111-1120.

Liu, W., Wang, X., Yang, X., &Xu, M. (2021).A Survey of Data Preprocessing Techniques for Machine Learning. In Proceedings of the 2021 3rd International Conference on

Liu, X., Zeng, Q., & Yu, C. (2021). What Drives Customer Satisfaction? A Literature Review of Antecedents and Consequences.International Journal of Hospitality Management, 94, 102890.

Marketing, 37(3), 347-367. doi: 10.1016/j.ijresmar.2019.11.004

Mohiuddin, M., Bhatti, A. I., Raza, G., &Minhas, M. R. (2021). Customer satisfaction prediction using machine learning algorithms: An empirical study of Pakistan's telecom sector. Journal of Business Research, 135, 17-31.

Mohiuddin, M., Bhatti, A. I., Raza, G., &Minhas, M. R. (2021). Customer satisfaction prediction using machine learning algorithms: An empirical study of Pakistan's telecom sector. Journal of Business Research, 135, 17-31

56

Nakandala, D., Hussain, F. K., & Chang, E. (2019).A comprehensive review on customer satisfaction prediction. International Journal of Information Management, 44, 48-60. <https://doi.org/10.1016/j.ijinfomgt.2018.10.015>

Oliver, R. L. (1977). A theory of communication and commitment effects in consumer satisfaction/dissatisfaction processes. Journal of Marketing Research, 4(3), 334-344.

Oliver, R. L. (1980). The development and application of a new instrument for measuring

Pandey, V., & Sharma, S. (2020).Predicting customer satisfaction in the banking sector using machine learning algorithms.International Journal of Information Management, 52, 102091.

Peppers, D., & Rogers, M. (1997). The One to One Future: Building Relationships One Customer at a Time. Currency Doubleday

Reichheld, F. F. (1993). Loyalty-based management. Harvard Business Review, 71(2), 64-73.

Reichheld, F. F. (2003). The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value. Harvard Business Review Press.

Rust, R. T., & Oliver, R. L. (1994). Service quality: Insights and opportunities from the frontier. Journal of the Academy of Marketing Science, 22(1), 1-19.

Samer M. Arqawi1 , Mohammed A. Abu Rumman2 , Eman Akef Zitawi3 , Basem S. Abunasser4 and Samy S. Abu-Naser5, Customer Satisfaction Prediction using Artificial Intelligence, 2022.

57

Shao, Y. (2021). A Study on the Factors Influencing Customer Satisfaction: Based on the Empirical Analysis of E-commerce Enterprises. International Journal of Online Marketing Research, 7(1), 18-30.

Shen, Y., Su, C., & Liu, Y. (2021).An intelligent customer feedback analysis system based on machine learning.Journal of Intelligent & Fuzzy Systems, 41(3), 3437-3449.

Wang, H., & Lo, H. P. (2021). Machine learning techniques for customer satisfaction prediction: A review and future directions. Expert Systems with Applications, 178,

Wang, X., Liu, Q., & Zhao, X. (2019).Predicting customer satisfaction with Airbnb services using machine learning algorithms. International Journal of Hospitality Management, 77, 246-257.

Weng, Q., Zhang, Y., & Ding, W. (2021).Exploring the key factors of online customer satisfaction in the sharing economy: Evidence from Airbnb.Telematics and Informatics, 61, 101582.

Westbrook, R. A. (1992). Value-Percept Disparity Theory: An alternative to the disconfirmation paradigm of consumer satisfaction. Journal of Consumer Research, 19(3), 305-314.

Yang, X., Zhang, X., & Chen, L. (2020).Hotel customer satisfaction prediction based on SVM.Journal of Ambient Intelligence and Humanized Computing, 11(6), 2433-2443.

58

Yang, Y., Fang, X., & Wei, C. (2021).A machine learning approach for customer satisfaction prediction. IEEE Access, 9, 128778-128790.

Zeithaml, V. A., Parasuraman, A., &Malhotra, A. (2008). Service quality delivered through websites: A critical review of extant knowledge and research agenda. Journal of the Academy of Marketing Science, 36(1), 1-19.

Zeithaml, V. A., Parasuraman, A., &Malhotra, A. (2008). Service quality delivered through websites: A critical review of extant knowledge and research agenda.

59.